CS 5350/6350: Machine Learining Spring 2018

Homework 3

Handed out: 26 March, 2018
Due date: 4 April, 2018

General Instructions

- You are welcome to talk to other members of the class about the homework. I am more concerned that you understand the underlying concepts. However, you should write down your own solution. Please keep the class collaboration policy in mind.
- Feel free discuss the homework with the instructor or the TAs.
- Your written solutions should be brief and clear. You need to show your work, not just the final answer, but you do *not* need to write it in gory detail. Your assignment should be **no more than 10 pages**. Every extra page will cost a point.
- Handwritten solutions will not be accepted.
- The homework is due by **midnight of the due date**. Please submit the homework on Canvas.
- Some questions are marked **For 6350 students**. Students who are registered for CS 6350 should do these questions. Of course, if you are registered for CS 5350, you are welcome to do the question too, but you will not get any credit for it.

1 Simple Conjunction Learner Analysis

1. [50 points] We have derived the PAC guarantee for the elimination algorithm learning simple conjunctions out of n input binary variables. Please use the PAC guarantee (in the lecture slides) to answer the following questions.

   (a) Suppose we have n=100. If we want to ensure with probability of at least 0.9, the generalization error of the learned hypothesis h is less than 0.1, how many training examples will we at least need?

   (b) Suppose n=100. If we want to ensure with probability of at least 0.99, the generalization error of the learned hypothesis h is less than 0.1, how many training examples will we at least need?

   (c) Suppose n=100. If we want to ensure with probability of at least 0.9, the generalization error of the learned hypothesis h is less than 0.01, how many training examples will we at least need?

(d) Suppose n=100. If we want to ensure with probability of at least 0.99, the generalization error of the learned hypothesis h is less than 0.01, how many training examples will we at least need?

(e) Suppose n=1000. If we want to ensure with probability of at least 0.99, the generalization error of the learned hypothesis h is less than 0.01, how many training examples will we at least need?

2. [10 points] Show that for $0 \leq x \leq 1$,

$$1 - x \leq e^{-x}.$$

2 Occam's Razor and PAC guarantee for consistent learners

We have derived the PAC guarantee for consistent learners (namely, the learners can produce a hypothesis which can 100% accurately classify the training data). The PAC guarantee is described as follows. Let H be the hypothesis space used by our algorithm. Let C be the concept class we want to apply our learning algorithm to search for a target function in C. We have shown that, with probability at least $1 - \delta$, a hypothesis $h \in H$ that is consistent with a training set of m examples will have generalization error $\mathrm{err}_D(h) < \epsilon$ if

$$m > \frac{1}{\epsilon}\big(\log(|H|) + \log\frac{1}{\delta}\big).$$

1. [20 points] Suppose we have two learning algorithms L_1 and L_2, which use hypothesis spaces H_1 and H_2 respectively. We know than H_1 is larger than H_2, i.e., $|H_1| > |H_2|$. For each target function in C, we assume both algorithms can find a hypothesis consist with the training data.

   (a) According to Occam's Razor principle, which learning algorithm's result hypothesis do you prefer? Why?

   (b) How is this principle reflected in our PAC guarantee? Please use the above inequality to explain why we will prefer the corresponding result hypothesis.

3 PAC Learnable Results

Let us check the PAC learnability for several concept classes. We assume the our learning algorithm can always result a hypothesis consistent with the training data. We assume the hypothesis space H is the same as the concept class C. Please determine whether the following concept classes are PAC learnable or not and prove your conclusions.

1. [10 points] General disjunctions out of n binary variables.

2. [10 points] m-of-n rules (Note that m is a fixed constant).

3. [10 points] Simple conjunctions out of n binary variables.

4. [10 points] k-CNF out of n binary variables.

5. [10 points] General boolean functions of n binary variables.